LA-UR-14-23202
Approved for public release; distribution is unlimited.

Title: Methodology for Sensitivity and Uncertainty-Based Criticality Safety Validation

Author(s): Kiedrowski, Brian C.

Intended for: Report

Issued: 2014-05-07

Methodology for Sensitivity and Uncertainty-Based Criticality Safety Validation

Brian C. Kiedrowski

1 Introduction

Monte Carlo radiation transport software is used extensively for determining the effective multiplication factor k of fissionable systems towards assessing the criticality safety of operations. Understanding the predictive capability of the Monte Carlo software and nuclear data libraries for calculations of specific applications is vital towards making these assessments. Recently, sensitivity and uncertainty (S/U) methods have been developed for this purpose. A computational tool has been developed that uses the S/U methods with MCNP6 [1] to select critical experiment benchmarks that have similar sources of computational bias, and extreme value theory is then used to determine suitable and conservative calculational margins. The S/U techniques are then used to perform nuclear data adjustments, which are then used to determine a suitable margin of subcriticality from cross section uncertainties.

This document gives an overview of the theory of S/U methods used for criticality safety validation. First an overview is given for using the sensitivity coefficients and nuclear covariance data towards determining neutronic similarity factors is then discussed. Next, the extreme value theory (EVT) to determine a calculational margin is given along with how the similarity factors are used in the analysis. Then, the Generalized Linear Least Squares (GLLS) method for performing data adjustment and how that method is used to reject benchmarks and compute adjusted nuclear data uncertainties in k are discussed. Last, a discussion of the margin of subcriticality is given along with how to apply of the GLLS data adjustment methods to quantify the effect of nuclear data variability on k.

2 Neutronic Similarity

The sensitivity coefficient to the effective multiplication factor k to some nuclear data x is defined as the relative change in k caused by some relative change in x:

$$S_{k,x} = \frac{\Delta k/k}{\Delta x/x}. \tag{1}$$

For an infinitesimal change in x, the sensitivity coefficient becomes

$$S_{k,x} = \frac{x}{k}\frac{dk}{dx}. \tag{2}$$

The magnitude of the sensitivity coefficient can be thought of as the impact that some piece of nuclear data has toward determining criticality, and its sign gives whether or not increasing the value of the data increases or decreases k.

The sensitivity coefficient can be estimated with Monte Carlo or deterministic methods using perturbation theory. MCNP6 uses the Iterated Fission Probability (IFP) method to compute the sensitivity coefficients in a continuous-energy Monte Carlo simulation [2], but other approaches are possible. The advantage to the IFP method is that, while it may not be the most efficient, it requires minimal user involvement and is therefore easy to automate. In either case, the similarity assessment is agnostic to the exact method used in the Monte Carlo, and the resulting sensitivity coefficients can be used interchangeably. The details of various approaches with multigroup cross sections and with continuous-energy physics are given in Refs. [2, 3, 4, 5, 6, 7, 8].

Although it cannot be shown rigorously, it is generally accepted that in most cases the predominant source of computational bias is because of uncertainties in the nuclear data. Since the sensitivity coefficients give which nuclear data matter most and the nuclear data uncertainties are the source of the bias, they, with the cross section covariance data to quantify the uncertainties, can be used to determine if two systems have similar biases.

The metric that is often used to assess a similar source of bias is called c_k, which is the correlation coefficient between two systems. The sensitivity coefficients are organized into a vector $\mathbf{S}$, whose elements represent various nuclides, reactions, and energy ranges. The nuclear covariance data is given as matrix $\mathbf{C}$ containing relative covariances. The sensitivity vectors and covariance matrix are consistently ordered. The correlation coefficient c_k can then be computed by

$$c_k = \frac{\mathbf{S}_1\mathbf{C}\mathbf{S}_2^T}{\sqrt{\mathbf{S}_1\mathbf{C}\mathbf{S}_1^T}\sqrt{\mathbf{S}_2\mathbf{C}\mathbf{S}_2^T}}. \tag{3}$$

Here the subscripts denote the two different systems being compared.

When $c_k = 1$, the two systems are neutronically identical and, in terms of nuclear data, have exactly the same source of computational bias. A value of $c_k = 0$ states that the two systems are completely dissimilar and therefore share no common nuclear-data driven source of computational bias. Negative c_k are theoretically possible indicating anti-similarity, but rarely observed and tend to be very small in practice. Empirically for low-enriched uranium (LEU) lattices, systems with c_k of greater than 0.9 or 0.95 have been shown to give reliable bias trends, and those with a c_k from 0.8 to 0.9 are marginally adequate. Unfortunately, the exact nature of the value of c_k to use as "similar enough" is system dependent; however, those with a higher c_k are more appropriate to use for validation than those with lower c_k.

3 Validation with Extreme Value Theory

Extreme Value Theory (EVT) may be applied to avoid the question of what value of c_k is similar enough for a given system. The application of EVT predicts probabilities of "worst case" occurrences and is conservative because increasing the sample size never decreases the predicted margins. The question then becomes whether or not the sample size is large enough. Answering this question rigorously for an arbitrary case is incredibly difficult, but some empirical numbers of sample sizes may be given that appear to work in practice.

EVT determines the probability that the maximum from a vector of random variables $\mathbf{X} = \{X_1, X_2, \ldots, X_N\}$ is less than some value x. Suppose each random variable X_j in $\mathbf{X}$ is independent and has a cumulative density function (CDF) F_j. The probability that x bounds the maximum X_j is given by the product of the CDF of F_j, or

$$F(x) = \prod_{j=1}^{N} F_j(x). \tag{4}$$

For criticality safety validation $F_j(x)$ is the random distribution function of the bias in k. Typically, the bias for an individual experiment is assumed to be normally distributed, and therefore the CDF is

$$F_j(x) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x - \beta_j}{\sqrt{2\sigma_j^2}}\right)\right], \tag{5}$$

where β_j is the computational bias determined from the difference in a calculation and a benchmark value and σ_j^2 is the variance of the bias, which is the sum in quadrature of the benchmark and Monte Carlo statistical uncertainties. For mathematical convenience, a positive β_j is defined to be non-conservative, i.e., the software predicts the benchmark value lower than it should. This change in the sign of the bias from what is typical needs to be consistently applied when determining the upper subcritical limit.

Each benchmark experiment can also be given a weight based upon how neutronically similar it is to the application of interest. If a benchmark has a weight w_j, the CDF F_j needs to be redefined such that as $x \to \infty$, $F(x) \to 1$. The corresponding probability density function can be thought of as having a Dirac delta function at $x = -\infty$ with weight $1 - w_j$ and a normal distribution centered about β_j with variance σ_j^2 having weight w_j. The weighted CDF is

$$F_j(x) = (1 - w_j) + \frac{w_j}{2}\left[1 + \text{erf}\left(\frac{x - \beta_j}{\sqrt{2\sigma_j^2}}\right)\right]. \tag{6}$$

For $w_j = 1$, the weighted CDF reduces to the unweighted CDF; for $w_j = 0$, the weighted CDF is unity for all x. From this point forward, F_j refers to the weighted CDF.

The benchmark weight w_j should represent the degree of similarity with an application system, which can be quantified by c_k. The choice of how the weight is assigned is rather arbitrary. Since the c_k is a linear correlation coefficient, a logical choice involves a linear interpolation of some sort. The scheme used is

$$w_j = \frac{c_{k,j} - c_{k,acc}}{c_{k,max} - c_{k,acc}}, \tag{7}$$

where $c_{k,j}$ is the similarity coefficient c_k for benchmark j, $c_{k,max}$ is the c_k of the most similar benchmark in the set, and $c_{k,acc}$ is the acceptance c_k (a lower bound of which benchmarks are used) to ensure an adequate total weight of the sample size w_{req}, such that

$$\sum_{j=1}^{N} w_j \geq w_{req}. \tag{8}$$

Determining the required weight is likewise arbitrary. The choice selected based upon empirical studies is

$$w_{req} = A + B\left(1 - c_{k,max}\right). \tag{9}$$

Here A is given as the minimum number of neutronically identical benchmarks or a set of non-identical benchmarks that add up to the same weight when there exists a benchmark that is neutronically identical $c_{k,max} = 1$. The parameter B is a penalty factor for not having a benchmark that is identical. The choice of B should be made such that for $c_{k,max} < 1$, more total benchmarks with $c_k = c_{k,max}$ (i.e., $w_j = 1$) are required than if $c_{k,max} = 1$.

The choices for the parameters A and B must be made empirically. The default values are $A = 25$ and $B = 100$. This implies that if there is a perfect benchmark in the set, 24 additional identical benchmarks are required or, more realistically, a number of benchmarks must add up to a total additional weight of 24. If $c_{k,max} = 0.95$, then 29 additional benchmarks are required with $c_{k,max} = 0.95$ or more with the statistical equivalent sum.

Once the sample benchmarks and sample weights w_j is determined, the EVT CDF $F(x)$ can be determined. The computational or calculational margin is found by finding the value of x where $F(x)$ equals the desired confidence. Typical choices are 0.95 or 0.99. For the latter choice of 0.99, this implies that there is a 99% chance that the value x bounds the worst case bias in k for the weighted sample set selected.

3.1 Discussion of Benchmark Selection and Extreme Value Theory

The premise of performing the validation is that there needs to be a sufficient number of benchmarks that are similar enough. In the case where there are insufficient benchmarks of high similarity, the method will automatically select benchmarks that are increasingly less suitable until the sample size criterion is met. The advantage of EVT is that adding a benchmark never decreases the calculational margin, so there is no concern about adding less relevant benchmarks from a safety point of view as the calculational margin will only become more conservative.

In practice, not having sufficient similar benchmarks available forces the method to dig more deeply into the suite, using a greater number of benchmarks, and increases the calculational margin to account for having less certainty. An example is Pu solutions with ENDF/B-VII.1 data. Near optimal moderation, there are a plethora of very similar benchmark experiments that calculate consistently and the predicted calculational margin is lower than for mor dilute or concentrated Pu solutions. In those regions, the validation does not have enough similar Pu solution benchmarks, so it tends to look at the Pu oxide, mixed oxide, and mixed solution benchmarks. The net effect again is to drive the calculational margin higher.

4 Generalized Linear Least Squares Data Adjustment

Typically the uncertainties in the nuclear data are the predominant source of computational bias. Given the sensitivity vector (first derivatives) and the covariance matrix for the benchmark experiments, it should be possible to perform an adjustment of the nuclear data that minimizes the bias within the bounds given by the covariance data. This can be used to find

benchmarks that are inconsistent with the assumption that their bias is driven by nuclear data uncertainties, and those can be rejected. The data adjustment process can also be used to adjust the covariances in the nuclear data as well given the benchmark experiments. These adjusted covariance data can be used to determine uncertainties in nuclear data.

A comprehensive coverview of the GLLS technique can be obtained in Refs. [9, 10] To summarize, the GLLS technique minimizes the χ^2 statistic, which is a quadratic function the sums the adjusted relative deviations of the calculation from benchmark and the proposed change in the nuclear data from its mean:

$$\chi^2 = [\Delta\mathbf{k}]^T \mathbf{C}_{kk} [\Delta\mathbf{k}] + [\Delta\mathbf{x}]^T \mathbf{C}_{xx} [\Delta\mathbf{x}] . \tag{10}$$

$\Delta\mathbf{k}$ represents the relative deviation in k for each benchmark after the adjustment predicted by the sensitivity coefficients, $\mathbf{C}_{kk}$ is the relative covariance matrix of the benchmark experiments, $\Delta\mathbf{x}$ is the relative deviation of the nuclear data from its mean, and $\mathbf{C}_{xx}$ is the relative covariance matrix for the nuclear data. With no adjustment, the $\Delta\mathbf{x}$ are all zero so the χ^2 is determined solely from the deviation in the benchmark experiments. As the data is adjusted, the deviation in the benchmark k may decrease, therefore decreasing the first term in the χ^2 equation, but this also increases the deviation in the nuclear data and therefore increases χ^2. Therefore, while it is almost always possible to find a data adjustment that reduces the experimental deviations to zero, given the very large number of degrees of freedom for adjusting the data, it is generally not possible to do this without increasing χ^2.

The goal is to find the data adjustment that minimizes χ^2 balancing the decrease from reducing the deviation of the benchmark experiments and the increase from growing the deviation of the nuclear data with its mean. There is a unique solution to this problem. First, define the deviation vector $\mathbf{d}$ as a vector of relative deviations in the calculation from benchmark experiments. The covariance matrix of the deviation vector for a set of benchmark experiments is

$$\mathbf{C}_{dd} = \mathbf{F}_{E/C}\mathbf{C}_{kk}\mathbf{F}_{E/C} + \mathbf{S}_{kx}\mathbf{C}_{xx}\mathbf{S}_{kx}. \tag{11}$$

$\mathbf{F}_{E/C}$ is a diagonal matrix containing the ratio of the benchmark k to calculated k and $\mathbf{S}_{kx}$ is a matrix where each row is the sensitivity vector for each benchmark experiment. Given these definitions, the minimum χ^2 is

$$\chi^2_{min} = \mathbf{d}^T\mathbf{C}_{dd}^{-1}\mathbf{d}, \tag{12}$$

where $\mathbf{C}_{dd}^{-1}$ is the inverse of the covariance matrix of the deviation vector.

4.1 Rejection of Inconsistent Benchmarks

The minimum χ^2 represents how well a regression model can fit the benchmark data through a nuclear data adjustment. Perfect agreement with the model would result in a χ^2_{min} divided by the number of degrees of freedom, i.e., the number of benchmarks, of unity. A χ^2_{min} of greater than one indicates that there are unexplained sources of bias other than the nuclear data uncertainties, or that the nuclear data and/or experimental covariances may be too small. Values of χ^2_{min} less than one indicates that the benchmark and/or nuclear data uncertainties are too large.

In practice, χ^2_{min} is going to be greater than unity, and the statistic may be used to reject inconsistent benchmarks until χ^2_{min} falls below a certain threshold. Exactly what the threshold should be is somewhat arbitrary, but a canonical value used for this application is that the χ^2_{min} divided by the number of benchmarks should be less than 1.2.

Two approaches are offered for the rejection of benchmarks. The first is the $\Delta\chi^2$ method, which is the most rigorous. While χ^2_{min} is greater than the threshold, the rejection routine recomputes for every benchmark the value of χ^2_{min} had that benchmark been rejected. The benchmark that it rejects is the one with that gives the greatest decrease in χ^2_{min}. If χ^2_{min} is greater than the threshold, the rejection routine proceeds anew and continues until χ^2_{min} is below the threshold.

Since this approach is very computationally costly for a large benchmark suite, the second more approximate approach is implemented. This is called the Iterative Diagonal χ^2 method. While the χ^2_{min} is less than the threshold, the diagonal χ^2 is computed for each benchmark:

$$\chi^2_{diag,j} = \mathbf{d}_j \mathbf{C}^{-1}_{dd}(j,j)\mathbf{d}_j. \tag{13}$$

Here $\mathbf{d}_j$ is the deviation in k from the jth benchmark, and $\mathbf{C}^{-1}_{dd}(j,j)$ is the corresponding diagonal element of the inverse of the deviation covariance matrix. The benchmark with the greatest χ^2_{diag} is rejected. As with the $\Delta\chi^2$ method, χ^2_{min} is recomputed and the rejection process continues until χ^2_{min} is recomputed. Note that for each rejection iteration, the inverse of the deviation covariance matrix $\mathbf{C}^{-1}_{dd}$ is recomputed. The Iterative Diagonal χ^2 method tends to reject a greater number of benchmarks than the $\Delta\chi^2$ method, but is significantly more computationally efficient.

Because the rejection of benchmarks is a function of the validation suite and independent of the applications being analyzed, the rejection of benchmarks is typically done once for the entire validation suite. The information about which benchmarks to reject, i.e., not include in the validation, is stored and read in when performing the validation with the EVT method.

4.2 Adjusted Nuclear Data Uncertainty

The nuclear data covariance matrix $\mathbf{C}_{xx}$ represents the prior uncertainties of the nuclear data given the differential measurements. In preparing a nuclear data evaluation, integral measurements, e.g., critical experiment benchmarks, are used to constrain the choices for cross sections as well, i.e., evaluators do not allow changes to nuclear data libraries that have too adverse an effect on the predictive performance of the critical experiment benchmarks. Because of this fact, there exists a dependency between the nuclear data and the critical experiment benchmarks, and the actual nuclear data uncertainties are therefore lower than the differential measurements alone would indicate.

The GLLS method, through its nuclear data adjustment, allows for an adjustment of the nuclear data covariances as well. The adjusted or residual covariance matrix post adjustment is found by

$$\mathbf{C}_{x'x'} = \mathbf{C}_{xx} - \left[\mathbf{C}_{xx}\mathbf{S}^T_{kx}\mathbf{C}^{-1}_{dd}\mathbf{S}_{kx}\mathbf{C}_{xx}\right]. \tag{14}$$

This residual covariance matrix $\mathbf{C}_{x'x'}$ may be used to determine an adjusted uncertainty in k because of the uncertainties in nuclear data. The sensitivity matrices $\mathbf{S}_{kx}$ are for the critical

experiment benchmarks, typically following a χ^2 rejection. The adjusted uncertainties in k for a set of applications may be found with the sandwich rule:

$$\mathbf{C}_{k'k'} = \mathbf{S}_{kx}\mathbf{C}_{x'x'}\mathbf{S}_{kx}^T. \tag{15}$$

Here $\mathbf{S}_{kx}$ is a matrix with rows as the sensitivity vectors for the applications, and $\mathbf{C}_{k'k'}$ is the covariance matrix for k of the applications. The square root of the diagonal elements of $\mathbf{C}_{k'k'}$ represent the adjusted relative uncertainties of k from the adjusted nuclear data covariances.

5 Margin of Subcriticality

In addition to the calculational margin, which is the bias plus bias uncertainty, an additional margin must be applied such that the analyst can ensure the process is actually subcritical. While this is ultimately the responsibility of the analyst as the details are inexorably linked with the process being analyzed, there are factors that the validation can address. First is the uncertainty in k because of variability in the cross sections, and second is the possible discrepancy in k from a calculation because of errors in transport and data processing software.

5.1 Variability in Cross Section Libraries

There are several modern cross section libraries available for performing criticality safety analysis. Because the preparation of a nuclear data evaluation involves expert judgment about which experimental results, theoretical models, and integral experiments to use, the results from one library almost always differ from another.

Typically, a new validation study selects a single cross section library that is recent. The choice of this is usually for non-technical reasons such as historical precedent and the geographic origin of the library relative to the facility (e.g., facilities in the US tend to use ENDF libraries, which are developed in the US; facilities in Europe tend to use JEFF libraries, which are developed in Europe). Because of this, there is no a priori reason to trust one set of experts over another, and different conclusions may result because of the choice of data library.

The traditional approach toward quantifying this is to swap out one nuclear data library with at least one other and to empirically determine the effect on the calculated k within the area of applicability. Doing this is not always easy from a workflow point of view, and there is always the question of why particular alternate libraries are selected over another.

The GLLS nuclear data adjustment methods offer an alternative way to determine the effect of this variability. In making an evaluation, both the cross section covariances and the critical experiment benchmarks are considered. The final nuclear data that gets released is usually chosen to be within their prior, differential measurement uncertainties and such that they do not have too adverse an impact on the critical experiment benchmarks. The reduced covariances following a GLLS nuclear data adjustment approximate the true uncertainty that an evaluator has available, in practice, for adjusting the nuclear data using their expert judgment.

Given this premise, the residual nuclear data covariance matrix can therefore be used via the sandwich rule to approximate the variation in k because of the variability in nuclear data libraries. Application of the sandwich rule implies an assumption of underlying multivariate normal distributions, and the standard rules of specifying confidence intervals apply. For example, at a 99% confidence interval, about 2.6 times the uncertainty in k from the variability in nuclear data can be applied towards determining a margin of subcriticality.

5.2 Undetected Errors from Transport and Nuclear Data Processing Software

Radiation transport software are a very complicated, and inevitably have errors that impact results in k. The same is true of the software used to process the nuclear data libraries from their native formats to those capable of being used by the transport software. Because of the unpredictable nature of software errors, these may not be detected as part of even an extensive validation.

As part of the process of determining an appropriate margin of subcriticality for this effect, a detection limit for the degree to which results would be impacted and to not be or have been noticed (else it would have already been fixed by the software developers) needs to be established. This needs to consider the typical experimental uncertainties in available benchmarks, the maturity of the software, the current amount of support, the current pedigree of the software quality assurance in the development, and the current and past amount of use for the application of interest. Software that is decades old and has currently and historically had thousands of users is more likely to have fewer errors than a newer software package with fewer than a hundred.

Determining the detection limit requires expert judgment by the software developers. For example, the MCNP software has a history dating back to before the 1970's, has been under continuous active development, and has over 10,000 users, of which a significant fraction perform criticality calculations of some sort. Considering that the best integral experiments have a benchmark uncertainty of 0.001 to 0.002, it is the expert opinion of the MCNP software developers that a suitable detection limit for k for criticality safety is 0.005. Errors of a larger magnitude would have already been addressed in such a mature software package.

The detection limit of 0.005 would also likely apply to KENO or SCALE as well, which has been used extensively for criticality safety analysis for decades. Less mature or used may require a detection limit of 0.01 to 0.03 depending on the maturity, level of use, support, and testing. In setting this detection limit, it is important to consult with the software developers to ascertain the degree to which they believe errors in k would have been detected.

Acknowledgments

Funding for this work was provided by the US DOE/NNSA Nuclear Criticality Safety Program and Advanced Scientific Computing Program. The author would like to thank Brad Rearden, Chris Perfetti, Don Mueller, and B.J. Marshall at Oak Ridge National Laboratory for very helpful discussions that made this work possible.

References

[1] MCNP6 Development Team, "Initial MCNP6 Release Overview", *Nucl. Technol.*, **180**, 298-315 (2012).

[2] B.C. Kiedrowski, F.B. Brown, "Adjoint-Based k-Eigenvalue Sensitivity Coefficients to Nuclear Data Using Continuous-Energy Monte Carlo," *Nucl. Sci. Eng.*, **174**, 227-224 (2013).

[3] C.M. Perfetti, W.R. Martin, B.T. Rearden, M.L. Williams, "Development of Continuous-Energy Eigenvalue Sensitivity Coefficient Calculation Methods in the Shift Mote Carlo Code," *Proc. PHYSOR 2012 – Advances in Reactor Physics*, Knoxville, TN, USA, Apr. 15-20 (2012).

[4] B.T. Rearden, "Perturbation Theory Eigenvalue Sensitivity Analysis with Monte Carlo Techniques," *Nucl. Sci. Eng.*, **146**, 367-382 (2004).

[5] B.T. Rearden, "TSUNAMI-3D: Control Module for Three-Dimensional Cross-Section Sensitivity and Uncertainty Analysis for Criticality," ORNL/TM-2005/39 Version 6, Oak Ridge National Laboratory (2009).

[6] K.F. Raskach, A.A. Blyskavka, "An Experience of Applying Iterated Fission Probability Method to Calculation of Effective Kinetics Parameters and keff Sensitivities with Monte Carlo," *Proc. PHYSOR 2010 – Advances in Reactor Physics to Power the Nuclear Renaissance*, Pittsburgh, Pennsylvania, USA, May 9-14 (2010).

[7] H.J. Shim, C.H. Kim, "Adjoint Sensitivity and Uncertainty Analyses in Monte Carlo Forward Calculations," *J. Nucl. Sci. Technol.*, **48**, 1453-1461 (2011).

[8] H.J. Park, H.J. Shim, C.H. KIM, "Uncertainty Quantification in Monte Carlo Criticality Calculations with ENDF/B-VII.1 Covariance Data," *Trans. Am. Nucl. Soc.*, **106**, 796-797 (2012).

[9] B.L. Broadhead, et. al., "Sensitivity- and Uncertainty-Based Criticality Safety Validation Techniques," *Nucl. Sci. Eng.*, **146**, 340-366 (2004).

[10] M.L. Williams, et. al., "TSURFER: An Adjustment Code to Determine Biases and Uncertainties in Nuclear System Responses by Consolidating Differential Data and Benchmark Integral Experiments," ORNL/TM-2005/39 Version 6, Oak Ridge National Laboratory (2009).